CS732: Data Visualisation Assignment 1 Report

Akanksha
DT2023001
Akanksha@iiitb.ac.in

Ketki Bhatia
DT2023007
Ketki.Bhatia@iiitb.ac.in

Niharika Suri
DT2023015
Niharika.Suri@iiitb.ac.in

***Abstract*—This project aims to investigate the geographic, demographic, and economic factors that influence credit card fraud patterns across the United States. Using a comprehensive dataset containing transaction details, fraud indicators, demographic information, and geographical coordinates, we analyzed patterns in fraudulent activity. Key variables included are fraud, transaction amounts, gender, age, and job categories. Various visualizations were employed to identify significant fraud trends across states and demographic segments, while outlier detection helped pinpoint common fraud-prone transaction amounts. The findings highlight that fraud is concentrated in economically active states such as New York and Pennsylvania, with certain job categories and transaction ranges being more vulnerable.**

I. Introduction

A. About the Dataset

The Credit Card Transactions Dataset provides detailed records of credit card transactions, including information about transaction times, amounts, and associated personal and merchant details. This dataset has over 1.85M rows. The dataset contains the following key attributes:

1) **Unnamed: 0:** This column is an index that would have been automatically created during data import, which does not hold significant value and may be ignored or dropped in further analysis.
2) **trans date trans time:** The timestamp of each transaction, which records the exact date and time when the transaction occurred.
3) **cc num:** A tokenized representation of the credit card number used for the transaction which ensures privacy while allowing transactions to be associated with specific customers.
4) **merchant:** The name or identifier of the merchant where the transaction took place. This field could be useful for identifying merchant-specific fraud trends.
5) **category:** The category of goods or services associated with the transaction (e.g., retail, groceries, electronics).
6) **amt:** The amount of money involved in the transaction.
7) **first:** The first name of the cardholder. This can be used for personalized analysis or segmentation of customer behaviors, though sensitive information should be anonymized.
8) **last:** The last name of the cardholder. Like the first name, it provides additional detail but should be anonymized in sensitive data handling.
9) **gender:** The gender of the cardholder, useful for demographic analysis of customer spending patterns and potentially identifying gender-based fraud trends.
10) **street:** The street address of the cardholder, which can be used for location-based analysis, though it should be handled carefully due to privacy concerns.
11) **city:** The city where the cardholder resides, useful for geographic analysis.
12) **state:** The state in which the cardholder resides. This allows for broader regional analysis of transaction trends or fraud detection within specific states.
13) **zip:** The ZIP code of the cardholder's address, which can be useful for geographic analysis.
14) **lat (Latitude):** The geographic latitude of the cardholder's residence. This field can be paired with the longitude to conduct geospatial analysis.
15) **long (Longitude):** The geographic longitude of the cardholder's residence. It is paired with latitude for accurate geolocation analysis.
16) **city pop:** The population of the cardholder's city.
17) **job:** The profession or occupation of the cardholder.
18) **job categories(created while data processing):** A broader classification of the job field or industry the cardholder belongs to. This helps categorise customers into general sectors for trend analysis.
19) **dob (Date of Birth):** The date of birth of the cardholder. This allows for age calculation, enabling demographic analysis based on the cardholder's age group.
20) **age(created while data processing):** The age of the cardholder at the time of the transaction.
21) **trans num:** A unique transaction number that identifies each transaction. This is important for tracking specific transactions and ensuring uniqueness in the dataset.
22) **unix time:** The transaction time represented in Unix timestamp format (seconds since January 1, 1970). This field provides a standardized time format for analysis.
23) **merch lat (Merchant Latitude):** The geographic latitude of the merchant's location.
24) **merch long (Merchant Longitude):** The geographic longitude of the merchant's location.
25) **is fraud:** A binary indicator that denotes whether the transaction is fraudulent (1) or legitimate (0).

Due to the extensive volume of data within the dataset, which spans the entire duration from January 1, 2019, to June 21, 2020, we focused our analysis on a subset of the dataset covering only the first six months of 2020. This decision was made to ensure that the analysis could be conducted within a reasonable timeframe. By concentrating on this six-month period, we aimed to maintain a manageable dataset size while still capturing relevant transaction trends and patterns for our analysis.

In the context of analysis, here are definitions for ”max states” and ”min states”:

- **Max States:** These are the 6 states with the highest number of credit card transactions, determined through our analysis using a standard deviation graph. This graph helped identify states with significantly higher transaction activity compared to the others.
- **Min States:** These are the 6 states with the lowest number of credit card transactions, identified using the standard deviation graph. This graph highlighted states with notably lower transaction activity relative to other states.

II. Methodology

A. Introduction

This represents a detailed approach to processing and analyzing a dataset using the Pandas library in Python, along with Tableau for data visualization. The primary focus is on fraud detection and job categorization, where various steps such as data inspection, cleaning, feature engineering, and classification were applied to prepare the data for analysis.

B. Workflow Overview

The dataset was processed in a series of steps to ensure proper cleaning, preparation, and analysis. The workflow began with an initial inspection of the dataset, followed by data extraction and cleaning. Feature engineering was then applied, particularly with the calculation of age, and job types were categorized for further analysis.

C. Data Processing Steps

1) Initial Data Inspection: The initial inspection of the dataset involved understanding its size and structure. The shape of the dataset was determined using the shape attribute, which provided insight into the number of rows and columns. To further analyse the data set’s structure, the info() function was employed. This function retrieved essential details such as column names, data types, and non-null counts, which helped identify any missing or inconsistent values in the data. Additionally, a statistical summary was generated using the describe() function. This summary provided key metrics such as the mean, median, minimum, maximum, and quartile values, giving a comprehensive understanding of the data distribution.

2) Data Subsetting and Cleaning: Once the initial inspection was complete, a subset of the dataset was extracted for further analysis. Specifically, the values from the 924,850th index onward were selected using the .iloc[] function. This step helped focus on the required portion of the data for subsequent analysis. The next step involved resetting the index using the reset index() function, ensuring continuity while discarding the old index. The last column of the dataset, which was deemed unnecessary, was dropped using the drop() function. To verify the removal, the head() function was applied, confirming the accuracy of the operation.

3) Handling Missing Data and Duplicates: To ensure the dataset was clean and ready for analysis, missing data was identified using the isnull().sum() function, which provided an overview of missing values across all columns. The dataset was also examined for duplicate rows using the duplicated() function. Upon calculation, it was determined that there were no duplicate records, so no further action was necessary to handle duplicates.

4) Feature Engineering: Age Calculation: The dataset included a dob (Date of Birth) column, which required transformation into a proper datetime format for consistency in date handling. The to datetime() function was employed to convert the dob column into a datetime format. A reference date of December 31, 2020, was established for calculating the age of each individual. The age was calculated by subtracting the dob from the reference date, and the result was divided by 365.25 to account for leap years. The calculated age values were then stored in a new age column.

5) Age Calculation in Tableau: In addition to Python-based calculations, the age of individuals was also computed in Tableau using a calculated field. To analyse the dataset by age, the dob field was converted into an age value by creating a calculated field. The formula used was based on the difference between the reference date (December 31, 2020) and the dob for each individual. This calculation ensured that each individual’s age was correctly determined, accounting for whether their birthday had occurred in the year of reference.

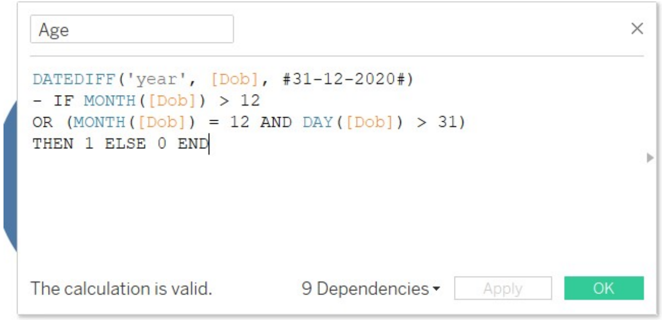
6) Job Categorisation: The dataset contained a variety of job types, and these were first extracted using the unique() function in Pandas to identify distinct job titles. After extracting the unique values, the job types were exported to Excel for further classification. In Excel, the unique job types were manually grouped into 10 broader categories based on relevance. To automate the categorization process, the VLOOKUP function was utilised. Specifically, the formula =VLOOKUP(Q:Q,’Job Dict.’!A:B, 2, FALSE) was applied to match each job type from the dataset with its corresponding category from a reference table and return the appropriate classification.

D. Calculated Fields of Tableau-

For creating customised visualisation depending upon the need we have created some calculated fields using the variables present in the dataset. The calculated variables are-

- Age- It is created using dob.
- Generation - It is created using age.
- Age in Range - It is also calculated using age
- Is Fraud(only 1) - It is calculated using Is Fraud.

Through this project, we aimed to test the hypothesis that regional factors significantly influence the distribution of credit card fraud across various states in the United States. Specifically, we focused on New York (NY), Pennsylvania (PA), Texas (TX), and California (CA) to assess whether the pro-

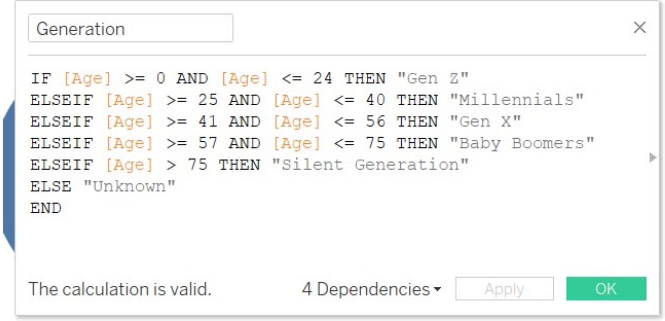


Age

```
DATEDIFF('year', [Dob], #31-12-2020#)
- IF MONTH([Dob]) > 12
OR (MONTH([Dob]) = 12 AND DAY([Dob]) > 31)
THEN 1 ELSE 0 END
```

The calculation is valid.

Fig. 1. Calculated Field of Age

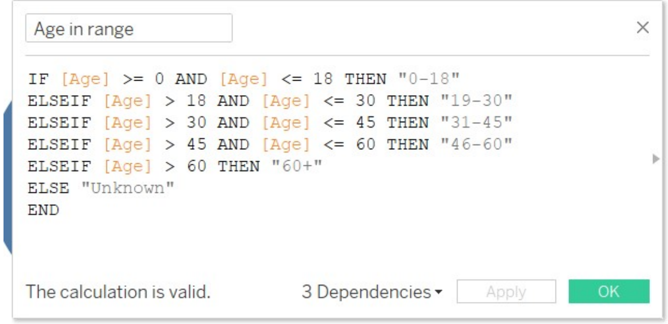


Generation

```
IF [Age] >= 0 AND [Age] <= 24 THEN "Gen Z"
ELSEIF [Age] >= 25 AND [Age] <= 40 THEN "Millennials"
ELSEIF [Age] >= 41 AND [Age] <= 56 THEN "Gen X"
ELSEIF [Age] >= 57 AND [Age] <= 75 THEN "Baby Boomers"
ELSEIF [Age] > 75 THEN "Silent Generation"
ELSE "Unknown"
END
```

The calculation is valid.

Fig. 2. Calculated Field of Generation

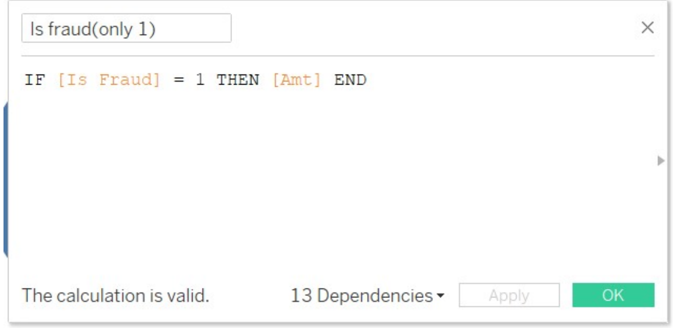


Age in range

```
IF [Age] >= 0 AND [Age] <= 18 THEN "0-18"
ELSEIF [Age] > 18 AND [Age] <= 30 THEN "19-30"
ELSEIF [Age] > 30 AND [Age] <= 45 THEN "31-45"
ELSEIF [Age] > 45 AND [Age] <= 60 THEN "46-60"
ELSEIF [Age] > 60 THEN "60+"
ELSE "Unknown"
END
```

The calculation is valid.

Fig. 3. Calculated Field of Age in Range



Is fraud(only 1)

```
IF [Is Fraud] = 1 THEN [Amt] END
```

The calculation is valid.

Fig. 4. Calculated Field of Is Fraud (only1)

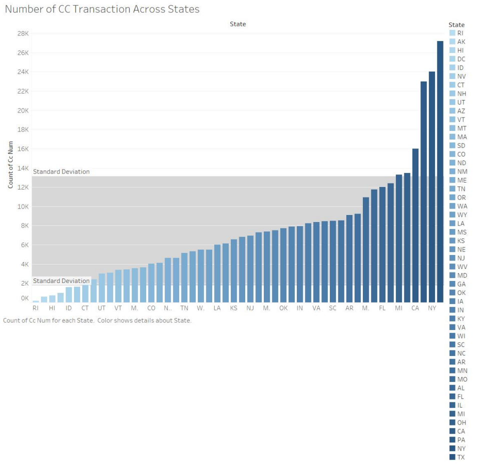


Fig. 5. It shows Number of CC Transaction Across States

portion of fraudulent transactions (is fraud) varies significantly between these regions.

III. Hypothesis

Null Hypothesis: There is no significant variation in the proportion of fraudulent transactions (is fraud) between the states of New York (NY), Pennsylvania (PA), Texas (TX), and California (CA).

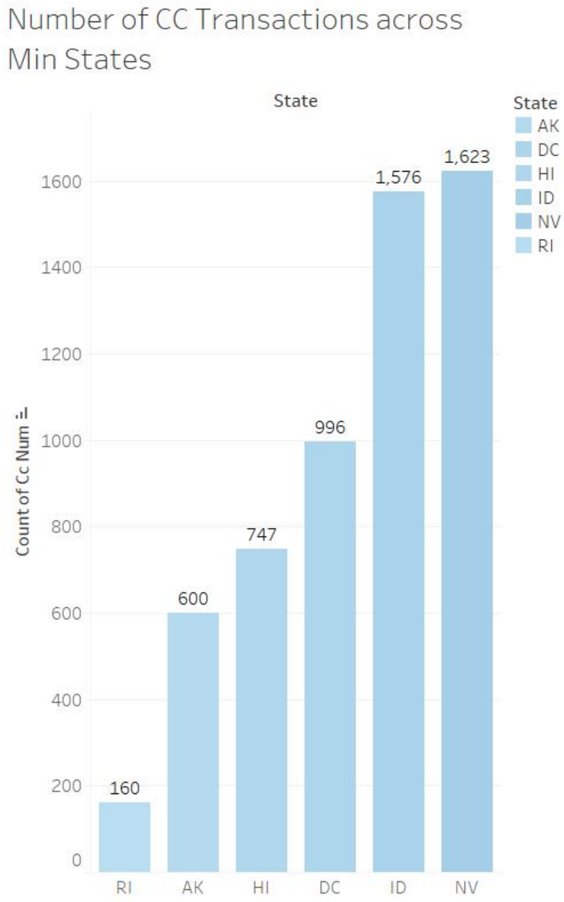
Alternate Hypothesis: There is a significant variation in the proportion of fraudulent transactions (is fraud) across the states of New York (NY), Pennsylvania (PA), Texas (TX), and California (CA), with certain states exhibiting higher or lower proportions of fraud due to varying regional factors.

IV. Analysis and Visualization

Credit Card Fraud in America- In the United States, credit and debit card transactions have become more prevalent than cash transactions. Unfortunately, this trend has also provided ample opportunities for digital criminals to commit credit card fraud. Last year, an alarming 52 million Americans fell victim to credit card fraud [1]. The Federal Trade Commission (FTC) received close to 390,000 reports of credit card fraud in 2021, making it one of the most prevalent types of fraud in the U.S. However, this figure may not fully capture the extent of the problem [1].

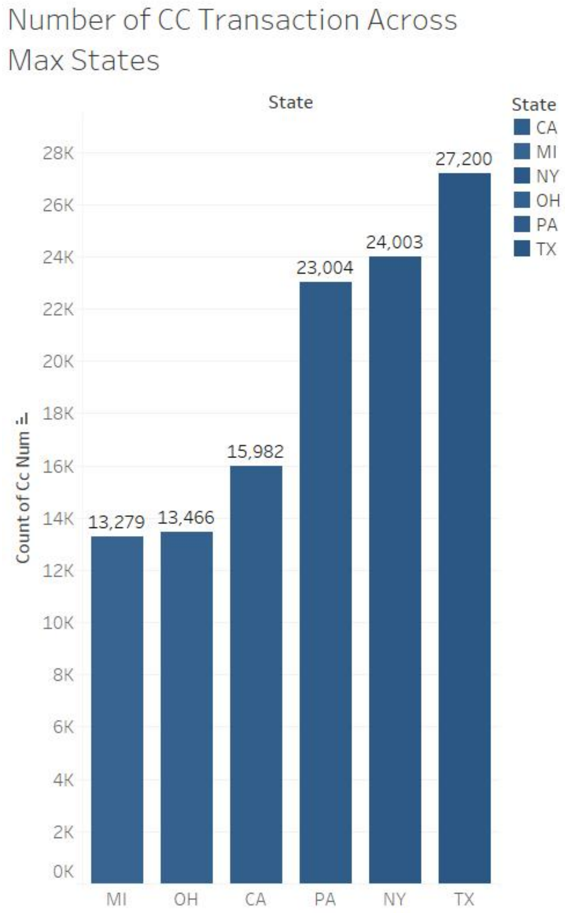
A. Exhibit 1: Number of CC Transaction Across States

Interpretation - We utilised the cc Num and state data to analyse the distribution of credit card transactions by creating a bar graph. In addition, We used standard deviation to identify any outliers. After conducting the analysis, we observed that Texas, New York, Pennsylvania, California, Ohio, and Michigan had the highest number of transactions(as shown in fig 3). It's noteworthy that four out of those six cities are present in the real data of the United States. On the other hand, Rhode Island, Alaska, Hawaii, Washington D.C., Idaho, and Nevada had the lowest number of transactions(as shown in fig 2). The intensity of the colour on the graph corresponds to the number



Count of Cc Num for each State. Color shows details about State. The marks are labeled by count of Cc Num. The view is filtered on State, which keeps 6 of 50 members.

Fig. 6. It shows Number of CC Transactions Across Min States



Count of Cc Num for each State. Color shows details about State. The marks are labeled by count of Cc Num. The view is filtered on State, which keeps 6 of 50 members.

Fig. 7. It shows Number of CC Transactions Across Max States

of transactions in each location.

The results show that Texas, New York, Pennsylvania, California, Ohio, and Michigan consistently recorded the highest number of transactions. This aligns with the fact that California, Texas, New York, and Pennsylvania were among the top 10 most populous states in the United States in 2021 [5]. High population sizes in these states, combined with their economic strength, likely contribute to the higher credit card activity observed. Notably, California, Texas, and New York each had a GDP exceeding 1 trillion dollar in 2021 and 2022, further reinforcing their financial activity [4].

Three of the states that appeared in our analysis California, Texas, and New York are among the most populous and have trillion-dollar economies, suggesting a strong correlation between population size, economic power, and credit card transaction volume. On the opposite end of the spectrum, Rhode Island, Alaska, Hawaii, Washington D.C., Idaho, and Nevada demonstrated the lowest number of credit card transactions.

Several factors contribute to this observation. Rhode Island, with its smaller population and more localized economy, naturally experiences fewer transactions [35]. Alaska and Hawaii, both geographically remote with lower population densities, face similar constraints. In Washington D.C., the economic focus is largely governmental, with a smaller consumer-driven economy relative to other states [4].

Idaho's rural economy and Nevada's reliance on tourism, particularly in specific regions like Las Vegas, could explain their lower overall transaction volumes during the analysis period [36 and 6]

From here onwards all the visualization have a factor of is fraud

B. Exhibit 2- Geographical Mapping of is Fraud (only1) Across United States

Interpretation-

Description: In this visualisation, we created a geographical map as it is suitable for visualising the spatial distribution of fraud occurrences by plotting average latitude and longitude.The states are labelled, and the sum of fraud occurrences (is fraud = 1) is represented by labels, with each point indicating a city. **Observations:** The map highlights distinct

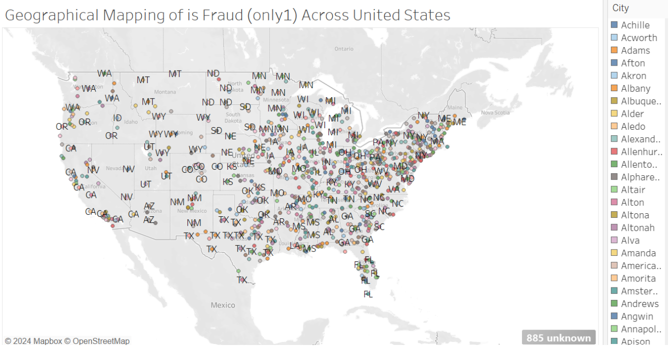
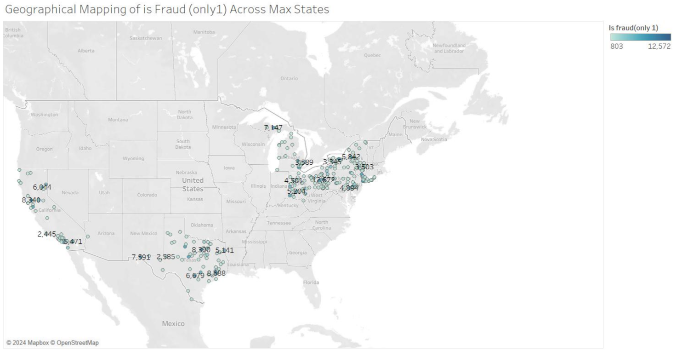


Fig. 8. Geographical Mapping of is Fraud (only1) Across United States



Map based on average of Long and average of Lat. Color shows sum of Is fraud(only 1). The marks are labeled by sum of Is fraud(only 1). Details are shown for City and State. The view is filtered on State, which keeps 6 of 50 members.

Fig. 9. Geographical Mapping of is Fraud (only 1) Maximum Transaction States

regional patterns in fraud activity, with varying concentrations across different parts of the U.S. The Northeast, Southeast, and Midwest show particularly high levels of fraud, while the Southwest and West have moderate to low concentrations.

In the Northeast, where there is a very high concentration of fraud, states like New York and Pennsylvania are key contributors. The region's financial hubs, such as New York City, along with densely populated urban areas, contribute to the high prevalence of fraud. The Northeast's reliance on financial services and technology-driven industries creates more opportunities for fraudulent activities, especially in sectors like banking, e-commerce, and online transactions [23].

The Southeast also exhibits a very high concentration of fraud, with states like Florida being particularly prone. This region has a large retiree population, which can be more vulnerable to certain types of fraud schemes, such as identity theft and financial scams targeting older individuals. Additionally, the Southeast's rapid urbanisation and growing reliance on digital transactions further expose its population to cyber fraud [39].

In the Midwest, where fraud activity is high but slightly lower than in the Northeast, states like Ohio and Michigan show significant fraud occurrences. These states have large manufacturing bases, but they are also evolving into tech and finance hubs [24]. With this industrial transition, there is an increasing number of digital transactions and online financial activities, which opens the door for more fraud incidents.

The Southwest shows a moderate-to-high concentration of fraud, particularly in Texas. Texas's rapidly growing economy, driven by industries such as energy, technology, and real estate, has seen an expansion of digital commerce. This economic growth, along with urbanisation and a diverse population, contributes to higher fraud rates as more people engage in online financial activities [12].

In the West, the concentration of fraud is low to moderate, with California being a notable exception. The state has a high population and a large tech economy, yet the lower fraud rates may be due to more advanced fraud detection systems and higher digital literacy compared to other regions. The presence of tech giants and cybersecurity initiatives in the West might also play a role in keeping fraud levels relatively contained.

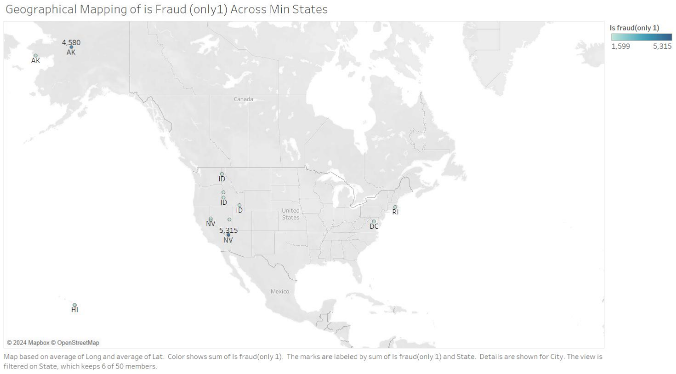
C. Exhibit 2.1- Geographical Mapping of is Fraud (only 1) Maximum Transaction States

Interpretation- We generated a geographical map focusing on states with high fraud occurrences.

Observations: In our analysis (Exhibit 2), we observed that fraud is concentrated predominantly along the Northeast and in the Midwest region. Notably, 4 out of the 6 high-transaction states Michigan (MI), Pennsylvania (PA), New York (NY), and Ohio (OH) are located in these regions, suggesting a geographic correlation between fraud activity and these areas. The concentration of fraud occurrences in states like Ohio (OH), Michigan (MI), Pennsylvania (PA), New York (NY), California (CA), and Texas (TX) can be attributed to specific regional and economic factors. In states like New York and Pennsylvania, major financial hubs such as New York City and Philadelphia drive high levels of credit card and online transactions, making these areas attractive targets for fraudsters. These cities are global centres for banking and commerce, where the sheer volume of financial activity increases the likelihood of fraud [38] [31].

In the Midwest, states like Ohio and Michigan also show high fraud occurrences. Ohio, with cities like Columbus and Cleveland, has a growing tech and financial services sector, but it may not have the advanced cybersecurity infrastructure seen in more tech-forward states, which could leave these systems more vulnerable to attacks. Similarly, Michigan's industrial history and the presence of large corporations and healthcare institutions create significant financial activity, which may increase the exposure to fraud [11] [16].

In California and Texas, which also exhibit significant fraud activity, the factors are somewhat different. California's major tech hubs, including Silicon Valley and Los Angeles, involve a high volume of online transactions and digital payments, which, while generally well-secured, present more opportunities for sophisticated fraud schemes like data breaches and phishing attacks [14]. Texas, particularly in cities like Houston and Dallas, experiences large-scale commerce in energy and finance, and rapid economic growth could lead to gaps in security as new systems and technologies are adopted, creating potential entry points for fraudsters [12].


Geographical Mapping of is Fraud (only1) Across Min States

Map based on average of Long and average of Lat. Color shows sum of Is fraud(only 1). The marks are labeled by sum of Is fraud(only 1) and State. Details are shown for City. The view is filtered on State, which keeps 6 of 50 members.

Fig. 10. Geographical Mapping of is Fraud (only1) Minimum Transaction States

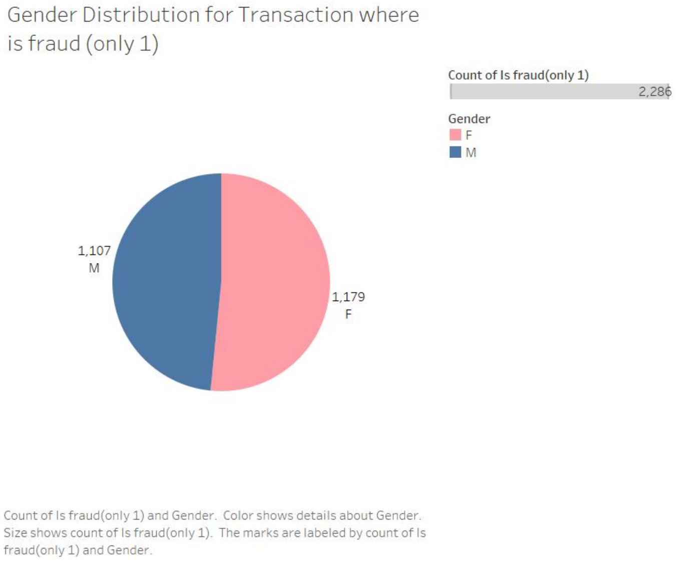
These states, despite their differing economic bases, share high levels of financial activity, which correlates with increased opportunities for credit card fraud, making them prime targets for malicious actors.

D. Exhibit 2.2- Geographical Mapping of is Fraud (only1) Minimum Transaction States

Interpretation- We generated a geographical map focusing on states with low fraud occurrences. **Observations:** We noticed that fraud cases are significantly lower in certain states. For example, states like DC, HI, ID, and RI have reported no fraud cases at all. On the other hand, NV (5,315) and AK (4,580) have reported cases of fraud. In Nevada, particularly in cities like Las Vegas, the high volume of tourism plays a significant role in increasing credit card fraud. Las Vegas is a major global destination, attracting millions of visitors each year who engage in frequent credit card transactions at hotels, casinos, restaurants, and entertainment venues [40]. The large number of tourists using unfamiliar networks, combined with the transient nature of transactions, creates opportunities for fraudsters to exploit vulnerabilities. Furthermore, the hospitality and service industries, which are heavily cashless, rely on high volumes of credit card transactions, increasing the chances of fraud in the state.

In Alaska, the reasons for credit card fraud are somewhat different. Despite its lower population density, Alaska experiences fraud due to its reliance on online transactions. The state's remote location means that many residents frequently shop online or use digital services for goods that are not locally available. This dependence on e-commerce increases the risk of online fraud, such as phishing and card-not-present scams. Additionally, Alaska's relatively smaller banking infrastructure compared to other states may lead to fewer fraud detection resources, potentially allowing fraudulent activities to go unnoticed for longer periods.

Notably, Washington D.C. (DC), Hawaii (HI), Idaho (ID), and Rhode Island (RI) did not report any fraud cases. Washington D.C., a region primarily focused on politics and government, has a unique economic structure with fewer consumer-driven transactions, which may explain the lack of reported fraud. The high level of governmental oversight and strict security measures might also contribute to reducing the likelihood of fraudulent activities [8].


Gender Distribution for Transaction where is fraud (only 1)

Count of Is fraud(only 1) and Gender. Color shows details about Gender. Size shows count of Is fraud(only 1). The marks are labeled by count of Is fraud(only 1) and Gender.

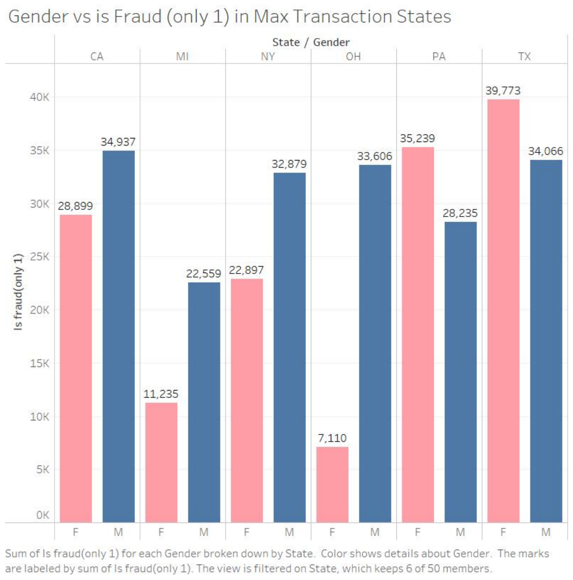
Fig. 11. Gender Distribution for Transaction where is Fraud (only1)

E. Exhibit 3 - Gender Distribution for Transaction where is Fraud (only1)

Interpretation-We created a pie chart to illustrate the breakdown of gender in fraudulent activities. In the chart, females are depicted in pink and males are shown in blue. Additionally, we included the count of fraud (only 1) alongside the gender in the labels for clarity. Our analysis revealed that fraud incidents involving females (1,179) were slightly more prevalent than those involving males (1,107).

Females are more frequently victims of credit card fraud partly due to their higher involvement in online shopping and digital financial activities, which increases exposure to scams (Delić, 2022). Additionally, a significant factor contributing to the elevated fraud rates among women is their vulnerability to romance scams. Action Fraud reports that over half of romance scam victims are women, with 50 percentage of victims being female compared to 39 percentage of male [41]. In these scams, fraudsters manipulate victims into sharing personal and financial information, including credit card details. The emotional manipulation in such scams makes women particularly susceptible, further driving up the incidence of credit card fraud targeting them.

Another possible explanation for these findings in our analysis is that our dataset consists of credit card transactions, and in the U.S., there are more female credit card holders than male credit card holders. This difference in the number of female and male credit card users may contribute to the observed correlations in our data, as the higher number of female credit card users could influence the patterns and trends we see [42].



Sum of Is fraud(only 1) for each Gender broken down by State. Color shows details about Gender. The marks are labeled by sum of Is fraud(only 1). The view is filtered on State, which keeps 6 of 50 members.

Fig. 12. It represents Gender vs is Fraud (only 1) in Max Transaction States

F. Exhibit 3.1 - Gender vs is Fraud (only 1) in Max Transaction States

Interpretation- We have analysed the correlation between gender and fraud in states with higher numbers of transactions. Using a bar chart, We compared the differences in fraud between males and females and with bars labelled with the count of fraud. By using bars labelled with the count of fraud, the chart effectively highlights differences in fraud occurrences between males and females.In the chart, the pink colour represents females while the blue colour represents males. We noticed that out of 6 states, in 4 states (CA, MI, OH, and NY), there were more cases of fraud involving males compared to females. Conversely, in 2 states (TX and PA), more fraud occurred with females compared to males.

In Michigan (MI), New York (NY), California (CA) and Ohio (OH), males were more frequently victims of fraud than females. This could be linked to the economic profiles and industries in these states, where males traditionally hold more roles in finance, business, and technology sectors, potentially exposing them to higher fraud risks through increased online transactions. Michigan, with its automotive and manufacturing base, and New York, a financial hub, are states where men may have greater financial engagement, which may elevate their risk of falling victim to fraud schemes [4].

In contrast, Texas (TX) and Pennsylvania (PA) reported higher fraud occurrences involving females. In these states, female consumers may have a more significant presence in online shopping or digital financial transactions, increasing their vulnerability. Texas, with its large, diverse economy and rapid urbanisation, could see higher fraud among females due to increased digital commerce [12]. Pennsylvania, a state with a growing retail and healthcare economy, might also have more women engaging in online services, making them frequent targets for fraud [13][4].

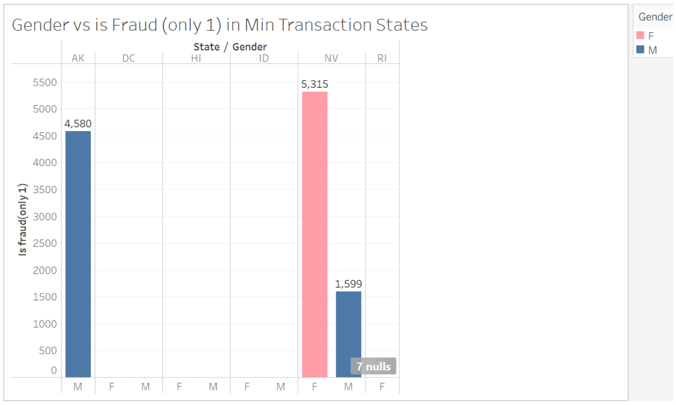


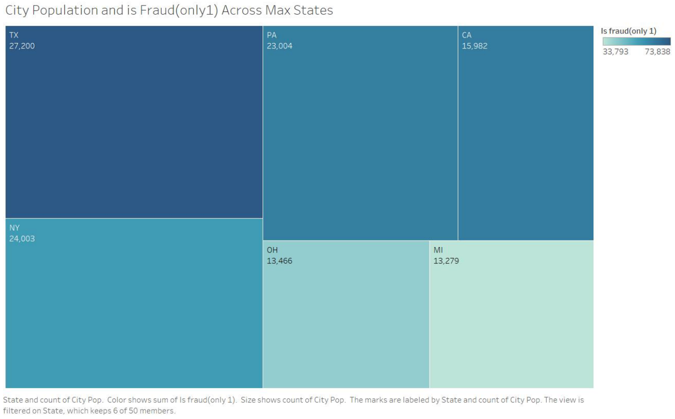
Fig. 13. It represents Gender vs is fraud (only 1) in Min states

G. Gender vs is fraud (only 1) in Min states

Interpretation- Similar to the previous graph(Fig-4) here we used a bar chart, in this visualisation we showcased gender and fraud for states with lower transaction volumes.The bar chart format highlights notable patterns, such as the absence of fraud cases in several states. Out of 6 states, specifically DC, HI, ID, and RI has no reported fraud cases. However, in AK, fraud was only associated with male transactions. In NV, fraud was observed in both male and female transactions. It's worth noting that, despite a low number of credit card transactions (1623), NV experienced a high number of fraud incidents: 5315 involving females and 1599 involving males. This visualisation illustrated the distribution of gender-based fraud across six states.On the other hand, Alaska (AK) reported fraud exclusively associated with male transactions. The remote geography and male-dominated industries, such as oil and fishing, in Alaska might expose men more frequently to high-value financial transactions, making them more susceptible to fraud [7].

Nevada (NV) stood out in the analysis. Despite having a relatively low number of credit card transactions (Exhibit 1), the state experienced an extraordinarily high number of fraud incidents: 5,315 involving females and 1,599 involving males. This is consistent with Nevada's ranking as having the highest rate of financial fraud per capita in the U.S [9]. The state's economy, particularly reliant on tourism, gambling, and entertainment, especially in Las Vegas, creates an environment full of opportunities for financial fraud. The heavy use of digital payments, coupled with the high volume of visitors and transactions, likely makes both residents and tourists vulnerable, with women being disproportionately targeted in this analysis [40].

(Exhibit 2.2) explains the reasons for no reported fraud cases. We noticed that states with lower ccNums have minimal data(as shown in Exhibit 3.1 and 3.2). Consequently, we have made the decision to focus exclusively on states with higher ccNums.



State and count of City Pop. Color shows sum of Is fraud(only 1). Size shows count of City Pop. The marks are labeled by State and count of City Pop. The view is filtered on State, which keeps 6 of 50 members.

Fig. 14. It represents City Population and is Fraud (only 1) Across Max States

H. Exhibit 4 - City Population and is Fraud (only 1) Across Max States

Interpretation- Description of the visualisation type used: In this visualisation, a heatmap was used to represent the relationship between the population of cities and the number of fraud occurrences across six states: TX, PA, CA, NY, OH, and MI. It allows for the examination of whether higher populations correlate with higher fraud occurrences, revealing insights into the distribution of fraud relative to city size.The city populations are counted and displayed on the heatmap, with the states labelled. The intensity of the colour corresponds to the number of fraud cases, where darker shades indicate higher fraud occurrences. **Observations:** From the heatmap, it is evident that Texas (TX) has the highest number of fraud cases (73,838), despite having a moderate population (27,200). Similarly, Pennsylvania (PA) and California (CA) both show a significant number of fraud occurrences (63,474 and 63,836 respectively) with relatively smaller populations (PA: 23,004; CA: 15,982). In contrast, states like Michigan (MI) and Ohio (OH) have smaller populations (13,279 and 13,466 respectively) and lower fraud counts (33,793 and 40,716). New York (NY), while having a higher population (24,003), has fewer fraud cases (55,777) compared to TX, PA, and CA.

This suggests that the number of fraud occurrences is not directly proportional to city population, indicating that other factors might contribute to fraud vulnerability in different states.

The heatmap highlights significant variations in fraud occurrences across six states; Texas (TX), Pennsylvania (PA), California (CA), New York (NY), Ohio (OH), and Michigan (MI) suggesting that factors beyond population size contribute to fraud rates. Here's a state-wise elaboration:

Texas stands out with the highest number of fraud cases (73,838) despite a moderate city population (27,200). Texas is home to several major cities like Houston, Dallas, and Austin, which are important financial, technological, and energy hubs [12]. These sectors involve large volumes of financial transactions,subsequently increasing the risk of fraud. The state's economic diversity ranging from oil and gas to tech startups may create more opportunities for different types of fraud [12]. Additionally, Texas has a high rate of online retail activity, which might be potentially contributing to more cyber fraud [12].

With 63,474 fraud occurrences, Pennsylvania has a significant number of cases, though its city population (23,004) is not as large as Texas. Philadelphia, a major urban centre, has a robust healthcare and financial services industry [7]. These sectors are often targets for fraud due to the sensitive personal information they handle . The relatively high rate of fraud may also be influenced by the state's older infrastructure and slower adoption of advanced security measures, making it more vulnerable to attacks like identity theft and financial fraud leading to increased statistics of credit card fraud.

California shows 63,836 fraud cases, despite having a smaller city population (15,982) compared to other states. As the most populous state in the U.S. and a global economic powerhouse, California's economy spans industries like technology, entertainment, and real estate, all of which involve substantial amounts of money [1]. Cities like Palo Alto, Los Angeles and San Francisco are centres for digital innovation but also hotspots for cybercrime, credit card fraud, and identity theft due to the high number of online transactions and the concentration of wealth [14]. California's large growth rate in population and the rise of fintech may also increase vulnerabilities in the financial system.

New York, with 55,777 fraud cases and a population of 24,003, sees fewer fraud occurrences compared to Texas and California, despite its status as a major financial centre. New York City is home to Wall Street and numerous financial institutions, which are often prime targets for fraud. However, stringent regulations and advanced fraud prevention technologies in the financial sector may mitigate some of these risks. The state's legal and regulatory frameworks might also play a role in reducing fraud rates compared to other states with similar economic activity.

New York, California, Texas, and Pennsylvania also ranked among the top 10 scammed states in the United States [27].

Ohio experiences 40,716 fraud cases with a relatively small population (13,466). While not an economic juggernaut like Texas or California, Ohio has a significant manufacturing and healthcare industry presence [16]. The healthcare sector, in particular, is susceptible to fraud, including insurance scams and identity theft.

Michigan reports 33,793 fraud cases, and the population is 13,279. Michigan's economy relies heavily on the auto industry, which may not present as many fraud opportunities as the finance or tech sectors in other states [11]. However, the state has seen an increase in cyber fraud, particularly related to auto loans and insurance scams [30].

According to the FBI Internet Report of 2021 Ohio and Michigan also ranked among the top 10 states to be affected by cyber crime [29].

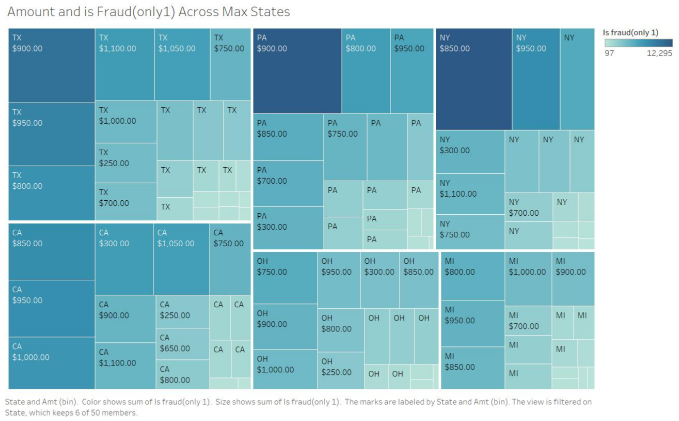


Fig. 15. It represents Amount and is Fraud (only 1) Across Max States

I. Exhibit 5- Amount and is Fraud (only 1) Across Max States

Interpretation- A heatmap was created where the colour intensity represents the number of fraud occurrences, so as to quickly spot which price ranges and states have higher or lower fraud rates. while the labels display the states and the transaction amounts in predefined bins. This visualisation provides a detailed breakdown of fraud cases at various price points across different states. In Tableau, we created and adjusted the bin size,for this we used the ”Create Bins” function to group the Amount column into intervals of 50 units to better visualise the distribution of transaction amounts in relation to fraud occurrences. **Observations:**

The heatmap reveals key insights into the distribution of fraud cases across different transaction amounts:

- **Texas (TX)** shows significant fraud activity, particularly in the 900 dollar range, with 10,189 fraud cases, followed by 8,699 cases at 800 dollar, and 7,523 cases in other categories.
- **Pennsylvania (PA)** also exhibits high fraud occurrences, with 12,052 cases at 900 dollar, and 6,602 at 800 dollar.
- **New York (NY)** shows prominent fraud counts, with 12,295 cases at 850 dollar and 7,813 at 950 dollar.
- **California (CA)** has notable fraud activity at 850 dollar (7,875 cases) and 950 dollar (7,798 cases), along with a significant number of fraud cases at the 1,000 dollar level (7,159 cases).
- **Ohio (OH)** shows a steady pattern, with 5,372 cases at 750 dollar, 4,674 at 900 dollar, and 4,117 at 1,000 dollar.
- **Michigan (MI)** displays a moderate amount of fraud cases, with 4,978 cases at 800 dollar, 4,817 at 950 dollar, and 4,069 at 1,000 dollar.

Overall, the analysis highlights that fraud tends to cluster around specific price ranges in different states, with certain states like Texas, Pennsylvania, and New York showing particularly high fraud activity in the 800–950 dollar range.

The concentration of fraud occurrences around the 800–950 dollar range in the heatmap suggests that these amounts may be strategically chosen by fraudsters. This range is likely optimal for evading detection, as transactions of this size are substantial enough to generate profit but not large enough to trigger automated security alerts. Many fraud detection systems flag unusually high or suspicious transactions, so mid-range amounts could fall below the threshold for heightened scrutiny.

Additionally, the 800–950 dollar range might align with common transaction amounts for certain goods or services, making these frauds blend in with legitimate purchases. Fraudsters may exploit typical consumer spending patterns, selecting amounts that are less likely to raise suspicion in both automated systems and manual reviews. The similarity in fraud patterns across states such as Texas, Pennsylvania, and New York indicates that fraudsters may employ similar strategies nationwide, taking advantage of systemic weaknesses in fraud prevention algorithms at these specific price points.

Finally, in states with a higher volume of financial activity, like Texas and New York, there may simply be more opportunities for fraud in this transaction range due to the larger number of high-value transactions. The clustering of fraud around these specific amounts likely reflects a combination of factors related to evasion tactics, consumer behaviour, and the structure of fraud detection mechanisms.

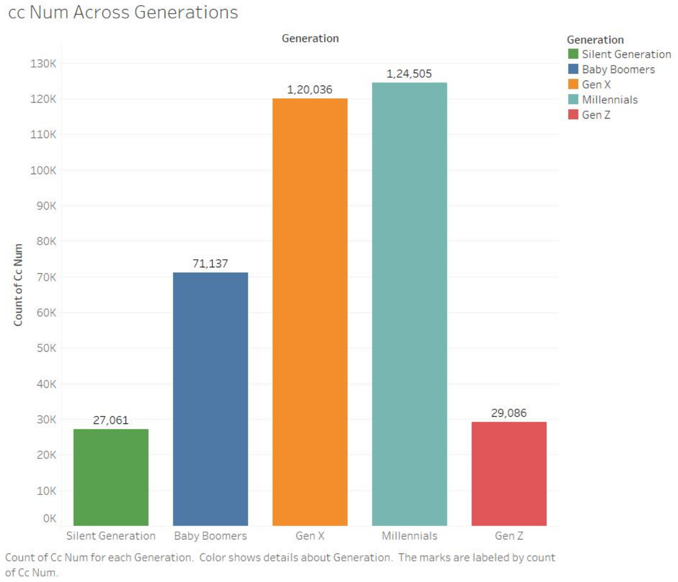
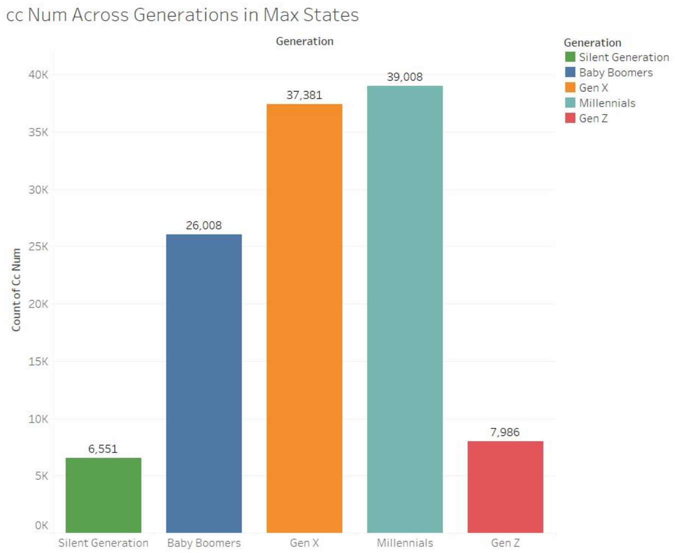


Fig. 16. It represents ccNum Across Generations

J. Exhibit 6 - ccNum Across Generations

Interpretation- We created a bar graph to illustrate the distribution of ccNum vs Generation by grouping individuals into different age brackets based on commonly accepted generational cohorts and with labels indicating the fraud value. By grouping individuals into generational cohorts and plotting the count of transactions (ccNum), the bar graph helps in effectively highlighting the differences in credit card usage among these age groups.We derived the age groups from the date of birth (dob) and categorized them as follows:


cc Num Across Generations in Max States

Count of Cc Num for each Generation. Color shows details about Generation. The marks are labeled by count of Cc Num. The data is filtered on State, which keeps 6 of 50 members.

Fig. 17. It represents ccNum Across Generations in Max States

- **Generation Z:** Individuals born between 1997 and 2012 (aged 8 to 23 in 2020)
- **Millennials (Generation Y):** Individuals born between 1981 and 1996 (aged 24 to 39 in 2020)
- **Generation X:** Individuals born between 1965 and 1980 (aged 40 to 55 in 2020)
- **Baby Boomers:** Individuals born between 1946 and 1964 (aged 56 to 74 in 2020)
- **Silent Generation:** Individuals born before 1946 (aged 75 and above in 2020)

We differentiated the generations using different colours in the graph and included the count of ccNum as labels to represent the numerical values.

Observations- The graph reveals that Millennials lead with the highest number of credit card transactions (124,505), closely followed by Gen X with 120,036 transactions, indicating that these two generations are the most active in credit card usage. When observed with Max States the same trend was observed as Millennials(39,008) have the highest number followed by Gen X(37,381).

The bar graph depicting the distribution of credit card numbers across different generations shows that Millennials lead with the highest number of credit card transactions, totalling 124,505. This is followed by Generation X with 120,036 transactions, Baby Boomers with 71,137 transactions, Generation Z with 29,086 transactions, and the Silent Generation with 27,061 transactions.

This distribution reflects the varying financial behaviours and life stages of each generation. Millennials, generally aged between their late 20s and early 40s, are at a point in their lives where they are likely to have greater financial responsibilities and opportunities, such as managing mortgages, car payments, and other significant expenses [18]. This phase often involves higher credit card usage for both everyday purchases and larger expenses. Additionally, Millennials are more likely to be comfortable with digital transactions and credit card management, contributing to their leading position in credit card transactions.

Generation X, aged between early 40's and late 50's, also shows a high level of credit card use, though slightly less than Millennials. This generation is often established in their careers and may have significant financial responsibilities, such as raising families and managing household expenses, which possibly drives their credit card usage [19]. Their high transaction count reflects their stable financial position and possibly higher spending capacity.

Baby Boomers exhibit a lower number of credit card transactions compared to the younger generations. This could be due to a more conservative approach to credit and spending, as well as different financial priorities and habits that were prevalent during their earlier years. This generation might also have a less frequent need for credit cards due to their established financial stability and reduced dependency on credit.

Generation Z, the youngest cohort, has the lowest transaction count. This is likely because they are still early in their financial lives, often in school or just starting their careers, and may have less disposable income and fewer credit cards [21]. Their lower transaction count could also be influenced by a more cautious approach to credit use compared to older generations.

The Silent Generation, being the oldest, shows the lowest transaction count. This may be due to a combination of factors, including a shift towards more conservative spending habits and potentially less engagement with credit cards as they move towards retirement [16]. Their financial needs and usage patterns are likely different from younger generations, contributing to their lower transaction figures.

K. Exhibit 7- is Fraud (only 1) Across Generations

Interpretation- We created a bar graph to illustrate the distribution of Generation where there is fraud only 1. By plotting the sum of fraud cases against each generation, the graph highlights which generational cohort is most affected by fraud.We have plotted generation on x-axis and sum of fraud on y-axis. Each generation is depicted in a different colour, with labels indicating the fraud value. The data shows that Gen X has the highest amount of fraud at 375,770, followed by Millennials at 359,401 when observed across United states but when compared with Max States it was found that Gen X (111,597) had highest amount of fraud but here it is followed by Baby boomers (86,953).

Generation X, as digital immigrants, had to adapt to new technology rather than being born into it. With established financial profiles, including high credit limits and significant assets, Generation X is a lucrative target for fraudsters. Their frequent use of both physical and online credit transactions increases their risk of exposure, and their susceptibility to fraud is further amplified by their relatively slower adoption of the latest cybersecurity measures. Millennials, being digital

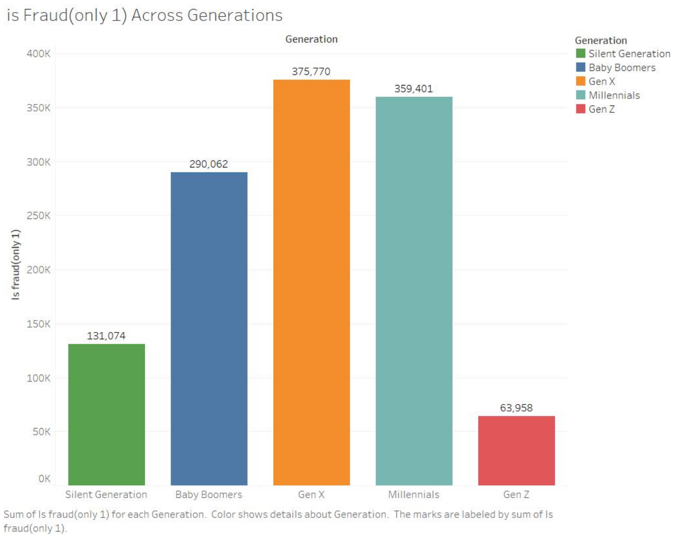


Fig. 18. It represents is Fraud (only 1) Across Generations

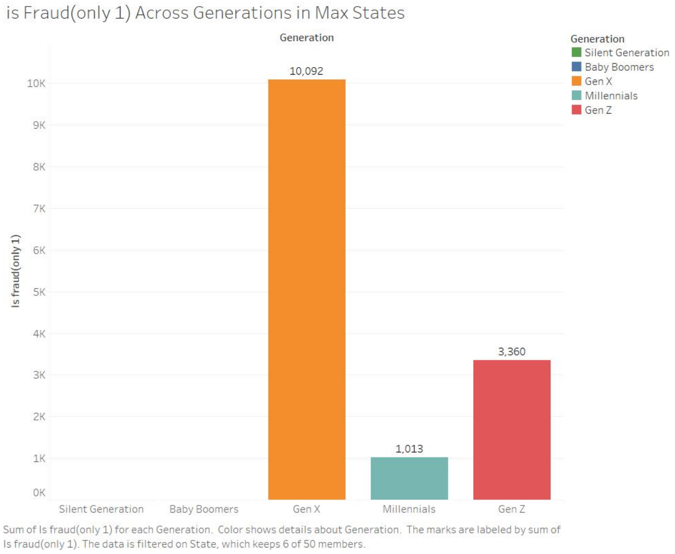


Fig. 19. It represents is Fraud (only 1) Across Generations in Max States

natives, rely heavily on online shopping, mobile banking, and digital payments.

Although they are comfortable with technology and can recognize common fraud risks, their high volume of digital transactions increases the likelihood of being targeted by cyber-attacks. Baby Boomers, as members of the late majority on the technology adoption curve, are slower to adopt the latest cybersecurity practices. Despite their lower overall digital engagement, their growing presence in online spaces exposes them to risks. Generation Z is highly familiar with technology and often early adopters of new digital tools. However, their comfort with technology can lead to complacency, making them vulnerable to sophisticated scams, particularly through social media and peer-to-peer payment platforms. The Silent Generation has the lowest number of fraud cases, largely due to their avoidance of extensive digital interaction, relying more on traditional banking methods like in-person banking and check writing.

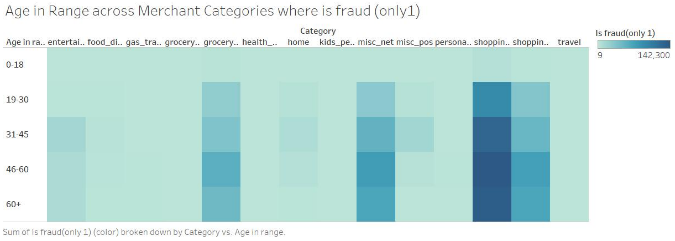


Fig. 20. Age in Range Across Merchant Categories where is fraud (only 1)

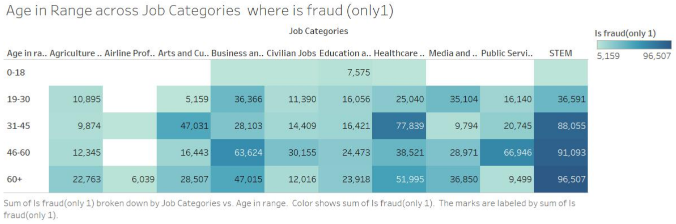
L. Exhibit 7.1- Age in Range Across Merchant Categories where is fraud (only 1)

Interpretation- In the dataset, we grouped individuals of different ages into specific ranges by applying a formula in the Calculated Field in Tableau. We then created a visualisation displaying age ranges, shopping categories, and fraud occurrences where fraud = 1. A heatmap was selected to emphasise the categories in which fraud occurred and the corresponding age ranges, with the colour intensity representing the number of fraud cases. We used this type of visualisation as heat map seems ideal for displaying the relationship between multiple variables, such as age ranges, shopping categories, and fraud occurrences.

This analysis revealed that the ”Shopping net” category was particularly susceptible to fraud, especially among individuals aged 19 to 60. The analysis of fraud occurrences based on age ranges and shopping categories revealed that the ”Shopping net” category was particularly vulnerable to fraud, especially among individuals aged 19 to 60. The heatmap visualisation clearly highlighted that this category stood out due to the significant number of fraud cases it represented.

Several factors could explain why ”Shopping net” emerged as a hotspot for fraud. First, online shopping has experienced exponential growth in recent years, particularly among the 19-60 (19-30, 31-45, 46-60) age group. This demographic, encompassing young adults to those in their middle years, is highly active in online transactions, relying on e-commerce for a wide range of goods and services. The convenience of online shopping has led to an increase in transactions, and with more people using the internet for purchases, the risk of encountering fraud has also risen. Fraudsters often target online shoppers because the digital environment offers multiple avenues for attack, such as phishing scams, data breaches, and fraudulent websites.

Younger individuals, especially those in the 19-30 range, may not have fully developed digital literacy or the necessary caution required for secure online transactions. As a result, they might fall prey to deceptive practices such as fake online stores or phishing attempts. On the other hand, individuals in the older segment of this range (45-60) may be targeted because they often engage in larger financial transactions



Age in Range across Job Categories where is fraud (only1)

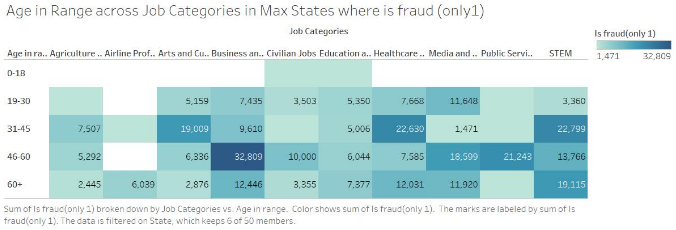
Job Categories

Age in ra..	Agriculture ..	Airline Prof..	Arts and Cu..	Business an..	Civilian Jobs	Education a..	Healthcare ..	Media and ..	Public Servi..	STEM
0-18						7,575				
19-30	10,895		5,159	36,366	11,390	16,056	25,040	35,104	16,140	36,591
31-45	9,874		47,031	28,103	14,409	16,421	77,839	9,794	20,745	88,055
46-60	12,345		16,443	63,624	30,155	24,473	38,521	28,971	66,946	91,093
60+	22,763	6,039	28,507	47,015	12,016	23,918	51,995	36,850	9,499	96,507

Is fraud(only 1)
5,159 96,507

Sum of Is fraud(only 1) broken down by Job Categories vs. Age in range. Color shows sum of Is fraud(only 1). The marks are labeled by sum of Is fraud(only 1).

Fig. 21. Age in Range across Job Categories in States where is fraud (only1)



Age in Range across Job Categories in Max States where is fraud (only1)

Job Categories

Age in ra..	Agriculture ..	Airline Prof..	Arts and Cu..	Business an..	Civilian Jobs	Education a..	Healthcare ..	Media and ..	Public Servi..	STEM
0-18										
19-30			5,159	7,435	3,503	5,350	7,668	11,648		3,360
31-45	7,507		19,009	9,610		5,006	22,630	1,471		22,799
46-60	5,292		6,336	32,809	10,000	6,044	7,585	18,599	21,243	13,766
60+	2,445	6,039	2,876	12,446	3,355	7,377	12,031	11,920		19,115

Is fraud(only 1)
1,471 32,809

Sum of Is fraud(only 1) broken down by Job Categories vs. Age in range. Color shows sum of Is fraud(only 1). The marks are labeled by sum of Is fraud(only 1). The data is filtered on State, which keeps 6 of 50 members.

Fig. 22. Age in Range across Job Categories in States where is fraud (only1) in Max States

online, whether for investments, home-related purchases, or travel.

Moreover, the age group (31-45) and (45-60) tends to have a higher disposable income and spending power, making them attractive targets for fraudsters. This age group often makes frequent and high-value purchases online, whether for necessities, luxury items, or subscription services, increasing their exposure to potential fraud risks.

M. Exhibit 7.2- Age in Range across Job Categories in States where is fraud (only1)

Interpretation- We utilised a highlight table for visualisation, as it is effective for showing the intersection of two categorical variables (age ranges and job categories) and their associated fraud occurrences, with age ranges placed in the rows and job categories in the columns. The intensity of the colour represented the number of fraud occurrences (where is fraud = 1), while also displaying the actual count of fraud cases. The results revealed that the STEM category had the highest number of fraud occurrences in the 60+ age group, followed by the 45-60 and 31-65 age ranges. Other job categories with significant fraud occurrences included Healthcare and Medicine, Public Services, and Law.

Next, we refined the visualisation by filtering the dataset to focus on the six states with the highest fraud occurrences (CA, PA, MI, TX, NY, OH). We examined whether the same job categories showed the highest fraud rates. The observations revealed a shift: the Business and Finance category now had the highest fraud occurrences in the 46-60 age range, while STEM and Healthcare and Medicine led in the 31-45 age range. Additionally, Public Services and Law saw higher fraud occurrences in the 46-60 age group.

The STEM category, particularly among individuals aged 60 and above, showed the highest fraud occurrences. This could be due to the fact that older professionals in STEM fields often have long-established careers with substantial incomes and savings. As they approach retirement, they may be less cautious with online transactions or more prone to phishing attacks, especially if they aren't as familiar with modern cybersecurity threats. Additionally, despite their technical backgrounds, they may not be as proactive in keeping up with evolving online security measures, making them susceptible to fraud.

The Business and Finance category had the highest fraud occurrences in the (46-60) age range. This demographic is typically at the peak of their financial and career stability, making them attractive targets for fraud. Individuals in this age group often handle larger financial transactions, manage multiple accounts, and engage in more complex financial activities, creating more opportunities for fraudsters to exploit vulnerabilities.

For individuals aged (31-45), STEM and Healthcare and Medicine job categories showed high fraud occurrences. This age group is likely to be highly active online and may engage in frequent online transactions, making them regular targets for cybercriminals. Healthcare professionals in particular have been targeted in recent years, with personal data often at risk due to the sensitive nature of their work and the large amount of personal information they handle. Additionally, professionals in STEM might be targeted due to the high-value nature of their work, despite their technical expertise.

The (46-60) age group also saw significant fraud occurrences in Public Services and Law. People in these professions are often targets for scams involving identity theft, as they may handle sensitive data or have access to government systems. Additionally, their high levels of trust within society and the nature of their work could make them less suspicious of fraudulent activities, especially if these scams appear to be official or government-related. As these individuals tend to be in trusted positions, fraudsters may also leverage this for more sophisticated schemes that prey on their professional roles.

V. Challenges and Limitations

During the course of this analysis, several challenges and limitations were encountered that could have influenced the findings. First, the dataset used primarily relied on transactional data, which may not fully capture all variables contributing to credit card fraud. The real-time variables, such as changes in economic conditions or technological advancements, were not accounted for, although they can significantly affect fraud trends.The geographical scope of the analysis also presents limitations. While the study covered specific states, fraud patterns likely vary across different regions and smaller jurisdictions, which were not fully explored. Another source of potential bias in this analysis arises from the categorization of jobs. The grouping of occupations was based on our interpretation and understanding, which may not align with industry-recognized classifications therefore certain nuances and distinctions between job roles might have been overlooked or oversimplified.

The analytical methods employed, such as visualisations and correlation analyses, while effective for identifying trends, do not account for deeper causal relationships, the use of advanced modelling techniques, such as machine learning-based fraud prediction models, could offer a more robust analysis. Another important limitation lies in the potential bias in the external sources used for validation. While U.S. websites were consulted to corroborate the analysis, online resources vary in reliability and may not provide up-to-date or region-specific data.

Together, these limitations highlight the complexity of analysing credit card fraud and emphasise the need for more comprehensive data, localised research, and advanced analytical methods in future studies.

VI. Conclusion

This analysis offers critical insights into the geographical, demographic, and transactional factors that shape credit card fraud patterns in the United States. The effective outlier detection in credit card fraud revealed that geographic and economic factors, rather than population size alone, drive fraud patterns. Gender, age, job categories, and specific transaction amounts also significantly influence fraud occurrences, with certain groups and transaction patterns showing higher vulnerability. Results of analysis shows that fraud is predominantly concentrated in economically significant eastern and midwestern states, such as New York, Pennsylvania, Michigan and Ohio, where high volumes of financial activity create increased exposure to fraudulent transactions. Notably, the analysis reveals no direct correlation between population size and fraud occurrences, suggesting that maybe economic dynamics, financial transaction volumes, and local consumer behaviours are more influential in shaping fraud risk. Demographic analysis highlights gender-based differences in fraud vulnerability, with men being more frequently affected in financially active states like New York and Ohio, while women show higher fraud involvement in states such as Texas and Pennsylvania. Furthermore, individuals aged 31 to 60, particularly those in high-income fields like STEM and Business and Finance, are disproportionately targeted, likely due to their financial engagement and larger transactional volumes.

The data also identifies a specific fraud pattern within transaction ranges of 800–950, suggesting that fraudsters intentionally exploit this range to avoid detection by conventional security systems. Moreover, the ”Shopping net” category emerges as a particularly high-risk area for fraud, reflecting the growing susceptibility of online shopping platforms to fraudulent activities, in nearly all age groups from 19 to 60.The results demonstrated a significant variation in fraud occurrences across the states of New York, Pennsylvania, Texas, and California, aligning with regional economic, demographic, and transaction volume factors. This supports the alternate hypothesis that certain states exhibit higher or lower proportions of fraud due to these regional factors. Thus, we reject the null hypothesis, confirming that there is indeed a meaningful difference in fraud distribution between the analysed states.”

In summary, the findings suggest that understanding fraud involves not just examining transaction volumes and population statistics but also considering regional economic characteristics, demographic behaviours, and strategic fraud tactics. Financial institutions should enhance their fraud detection mechanisms by addressing these vulnerabilities. By doing so, they can more effectively mitigate the growing threat of credit card fraud.

VII. Acknowledgement

- **Ketki Bhatia and Niharika Suri (Dataset Preparation and Data Processing):** They collaboratively worked on cleaning, organising, and structuring the dataset to ensure consistency and accuracy. They handled missing values, corrected data inconsistencies, and pre-processed the dataset to make it suitable for analysis.
- **Akanksha (Calculated Fields in Tableau):** Akanksha developed and implemented calculated fields in Tableau to derive additional insights, such as generating age from date of birth and categorising transaction amounts. These fields provided essential metrics for the visualisations.
- **Akanksha and Ketki Bhatia (Visualisation Exhibit 1-5):** They created the initial set of visualisations (Exhibits 1-5), including bar charts, heatmaps. They visualised patterns such as fraud occurrence by state, gender-based fraud distribution, and transaction amount analysis.
- **Niharika Suri and Akanksha (Visualisation Exhibit 6-7):** They collaborated on Exhibits 6 and 7, which involved deeper visual analysis, focusing on specific states and gender-based fraud comparisons. They also employed specialised techniques such as state-level fraud mapping.
- **Niharika Suri (Research about notable characteristics and unique features of various U.S. states):** Niharika conducted research on the economic profiles, industries, and regional vulnerabilities of key U.S. states to provide context for the fraud trends observed. This research helped explain why certain states, like Texas and Pennsylvania, had high fraud occurrences.
- **Ketki Bhatia (Research about various generations and their characteristics):** Ketki performed research on generational behaviours, focusing on how different age groups engage with digital platforms and financial systems. This research was used to understand demographic factors in fraud vulnerability, particularly how older generations or younger tech-savvy individuals might be targeted differently by fraudsters.

References

[1] B. Cruz, “52 million Americans experienced credit card fraud last year,” *Security.org*, Jul. 26, 2024. [Online]. Available: https://www.security.org/digital-safety/credit-card-fraud-report/: :text=60

[2] J. Egan, “Credit card fraud statistics,” *Bankrate*, Jan. 12, 2023. [Online]. Available: https://www.bankrate.com/credit-cards/news/credit-card-fraud-statistics/.

[3] D. Carlin, “All 50 US states ranked by GDP [Report 2024],” *USA by Numbers*, Jan. 14, 2023. [Online]. Available: https://usabynumbers.com/states-ranked-by-gdp/.

[4] USAFacts, “How does gross domestic product differ by state?,” *USAFacts*, Dec. 05, 2023. [Online]. Available: https://usafacts.org/articles/how-does-gdp-differ-by-state/.

[5] GlobalData, “Most populated states in the United States in 2021,” [Online]. Available: https://www.globaldata.com/data-insights/macroeconomic/most-populated-states-in-the-us/.

[6] Gigafact, “Fact Brief: Is Nevada’s economy heavily dependent on tourism?” Nov. 10, 2020. [Online]. Available: https://gigafact.org/fact-briefs/nevadas-economy-heavily-dependent-tourism.

[7] M. M. Miller and D. Lynch, “Alaska — History, flag, Maps, weather, cities, and Facts,” *Encyclopedia Britannica*, Sep. 13, 2024. [Online]. Available: https://www.britannica.com/place/Alaska.

[8] H. J. Critchfield, E. Clark, A. Augustyn, and G. L. McNamee, “Washington — State Capital, Map, History, cities, and Facts,” *Encyclopedia Britannica*, Jul. 26, 1999. [Online]. Available: https://www.britannica.com/place/Washington-state.

[9] G. America, “Nevada ranked as having highest rate of financial fraud per capita,” [Online]. Available: https://gamingamerica.com/news/9247/nevada-ranked-as-having-highest-rate-of-financial-fraud-per-capita.

[10] Oberlo, “Top US payment methods (2023–2027),” [Online]. Available: https://www.oberlo.com/statistics/top-us-payment-methods.

[11] J. Hallinen, “STEM — Description, Development, and Facts,” *Encyclopedia Britannica*, Sep. 06, 2024. [Online]. Available: https://www.britannica.com/topic/STEM-education/STEM-educationref330962.

[12] R. A. Wooster, D. C. Reddick, and G. L. McNamee, “Texas — Map, population, History, and Facts,” *Encyclopedia Britannica*, Sep. 14, 2024. [Online]. Available: https://www.britannica.com/place/Texas-state.

[13] C. L. Thompson and E. W. Miller, “Pennsylvania — Capital, Population, Map, Flag, Facts, and History,” *Encyclopedia Britannica*, Sep. 14, 2024. [Online]. Available: https://www.britannica.com/place/Pennsylvania-state.

[14] N. Morgan and G. L. McNamee, “California — Flag, Facts, Maps, Capital, Cities, and Destinations,” *Encyclopedia Britannica*, Sep. 14, 2024. [Online]. Available: https://www.britannica.com/place/California-state.

[15] P. J. Scudiere and A. K. Campbell, “New York — Capital, map, population, history, and facts,” *Encyclopedia Britannica*, Sep. 13, 2024. [Online]. Available: https://www.britannica.com/place/New-York-state.

[16] F. R. Aumann, G. W. Knepper, and J. Wallenfeldt, “Ohio — History, capital, population, map, and Facts,” *Encyclopedia Britannica*, Sep. 13, 2024. [Online]. Available: https://www.britannica.com/place/Ohio-state.

[17] R. J. Hathaway, S. Glazer, and R. J. Schaetzl, “Michigan — Capital, Map, Population, History, and Facts,” *Encyclopedia Britannica*, Sep. 14, 2024. [Online]. Available: https://www.britannica.com/place/Michigan.

[18] A. Zelazko, “Millennial — Definition, characteristics, age range, and birth Years,” *Encyclopedia Britannica*, Sep. 04, 2024. [Online]. Available: https://www.britannica.com/topic/millennialref356992.

[19] A. McKenna, “Generation X — Origin, Years, Characteristics, and Facts,” *Encyclopedia Britannica*, Sep. 04, 2024. [Online]. Available: https://www.britannica.com/topic/Generation-Xref356273.

[20] P. Bump, “Baby boomer — Definition, Age Range, and Societal and Economic Impact,” *Encyclopedia Britannica*, Sep. 13, 2024. [Online]. Available: https://www.britannica.com/topic/baby-boomers.

[21] A. Eldridge, “Gen Z — Years, Age Range, Meaning, and Characteristics,” *Encyclopedia Britannica*, Sep. 13, 2024. [Online]. Available: https://www.britannica.com/topic/Generation-Z.

[22] J. Wallenfeldt, “Silent Generation — Years, characteristics, and name Meaning,” *Encyclopedia Britannica*, Aug. 16, 2024. [Online]. Available: https://www.britannica.com/topic/Silent-Generation.

[23] The Editors of Encyclopaedia Britannica, “Eastern Seaboard — Map, Region, and Facts,” *Encyclopedia Britannica*, Sep. 12, 2024. [Online]. Available: https://www.britannica.com/place/Eastern-Seaboard.

[24] The Editors of Encyclopaedia Britannica, “Midwest — History, States, Map, culture, and facts,” *Encyclopedia Britannica*, Sep. 14, 2024. [Online]. Available: https://www.britannica.com/place/Midwest.

[25] S. Burga, “Why Gen Z is surprisingly susceptible to financial scams,” *TIME*, Feb. 24, 2024. [Online]. Available: https://time.com/6802011/gen-z-financial-scams-fraud/.

[26] M. Michaels, “All 50 states ranked for identity theft and credit card fraud, from most at risk to the least,” *Business Insider India*, Feb. 15, 2018. [Online]. Available: https://www.businessinsider.in/all-50-states-ranked-for-identity-theft-and-credit-card-fraud-from-most-at-risk-to-the-least/articleshow/62937494.cms.

[27] N. Campisi, “The 10 most scammed states in America,” *Forbes Advisor*, Aug. 04, 2023. [Online]. Available: https://www.forbes.com/advisor/personal-finance/most-scammed-states/.

[28] B. Cruz, “52 million Americans experienced credit card fraud last year,” *Security.org*, Jul. 26, 2024. [Online]. Available: https://www.security.org/digital-safety/credit-card-fraud-report/: :text=60

[29] J. Egan, “Credit card fraud statistics,” *Bankrate*, Jan. 12, 2023. [Online]. Available: https://www.bankrate.com/credit-cards/news/credit-card-fraud-statistics/.

[30] MSN, “Ohio ranks among states most affected by internet crimes according to FBI,” *MSN*, Sep. 14, 2024. [Online]. Available: https://www.msn.com/en-us/news/us/ohio-ranks-among-states-most-affected-by-internet-crimes-according-to-fbi/ar-AA1qoOBP.

[31] D. Scofield and T. Carloss, “Ohio listed no. 7 for number of victims of cybercrime country-wide in 2021,” *News 5 Cleveland WEWS*, Apr. 12, 2022. [Online]. Available: https://www.news5cleveland.com/news/local-news/ohio-listed-no-7-for-number-of-victims-of-cybercrime-country-wide-in-2021.

[32] S. Taber, “Internet Crime Report - Michigan SBDC,” *Michigan SBDC*, Mar. 15, 2023. [Online]. Available: https://michigansbdc.org/cybersecurity/internet-crime-report/.

[33] M. S. Magda, J. B. B. Trussell, and S. K. Stevens, “Philadelphia — History, map, population, and Facts,” *Encyclopedia Britannica*, Sep. 13, 2024. [Online]. Available: https://www.britannica.com/place/Philadelphia.

[34] P. R. Duis and C. Schallhorn, “Chicago — History, population, map, and facts,” *Encyclopedia Britannica*, Sep. 12, 2024. [Online]. Available: https://www.britannica.com/place/Chicago.

[35] J. S. Lemons, “Rhode Island — Map, population, history, beaches, and Facts,” *Encyclopedia Britannica*, Sep. 15, 2024. [Online]. Available: https://www.britannica.com/place/Rhode-Island-state.

[36] B. A. Martin and G. L. McNamee, “Idaho — History, Economy, People, and Facts,” *Encyclopedia Britannica*, Sep. 15, 2024. [Online]. Available: https://www.britannica.com/place/Idaho.

[37] C. Lamothe, “5 reasons women are more likely to be targeted for financial theft scams,” *GOBankingRates*, Sep. 14, 2024. [Online]. Available: https://www.gobankingrates.com/money/financial-planning/reasons-women-are-more-likely-to-be-targeted-for-financial-theft-scams/.

[38] G. Lankevich, “New York City — Layout, map, economy, culture, facts, and History,” *Encyclopedia Britannica*, Sep. 15, 2024. [Online]. Available: https://www.britannica.com/place/New-York-City.

[39] The Editors of Encyclopaedia Britannica, “The South — Definition, States, Map, and History,” *Encyclopedia Britannica*, Sep. 15, 2024. [Online]. Available: https://www.britannica.com/place/the-South-region.

[40] G. L. McNamee, “Las Vegas — History, Layout, Population, Map, Economy, and Facts,” *Encyclopedia Britannica*, Sep. 15, 2024. [Online]. Available: https://www.britannica.com/place/Las-Vegas-Nevada.

[41] D. Delić, “Are women at more risk of online scams? The latest statistics in 2022,” *ProPrivacy.com*, Jul. 06, 2022. [Online]. Available: https://proprivacy.com/blog/women-and-online-scams-latest-statistics-2022.

[42] C. Rodriguez, “Credit Card Ownership Statistics and Facts – By Income, Credit Score, Education Level and More [2024 Data Study],” *UpgradedPoints.com*, Sep. 13, 2024. [Online]. Available: https://upgradedpoints.com/credit-cards/credit-card-ownership-statistics/.